Academic careers in Computer Science: Continuance and transience of lifetime co-authorships

Guillaume Cabanac, Gilles Hubert, Béatrice Milard

DOI 10.1007/s11192-014-1426-0

Academic careers in Computer Science: Continuance and transience of lifetime co-authorships

Guillaume Cabanac · **Gilles Hubert** · **Béatrice Milard**

Abstract Scholarly publications reify fruitful collaborations between co-authors. A branch of research in the *Science Studies* focuses on analyzing the co-authorship networks of established scientists. Such studies tell us about how their collaborations developed through their careers. This paper updates previous work by reporting a transversal and a longitudinal studies spanning the lifelong careers of a cohort of researchers from the DBLP bibliographic database. We mined 3,860 researchers' publication records to study the evolution patterns of their co-authorships. Two features of co-authors were considered: 1) their expertise, and 2) the history of their partnerships with the sampled researchers. Our findings reveal the ephemeral nature of most collaborations: 70% of the new co-authors were only one-shot partners since they did not appear to collaborate on any further publications. Overall, researchers consistently extended their co-authorships 1) by steadily enrolling beginning researchers (i.e., people who had never published before), and 2) by increasingly working with confirmed researchers with whom they already collaborated.

Keywords Co-authorship networks · Research collaboration · Research careers · Cohort study · Transversal study · Longitudinal study · Partnership ability

Introduction

Since the seminal work of Beaver and Rosen (1978), the rationale behind collaborations between researchers has been a recurring matter of interest. It has been found that collaborations are facilitated by technological advances (Katz & Martin, 1997), by geographic proximity (Landry & Amara, 1998), and more generally by homophily (i.e., similarity of research topics or status, see (Evans, Lambiotte, & Panzarasa, 2011)). However, differences across disciplines have been highlighted (Abramo, D'Angelo, & Di Costa, 2009), as well as between diverse types of research groups (Stvilia et al., 2011). Overall, it appears that

G. Cabanac · G. Hubert
Computer Science Department, University of Toulouse, IRIT UMR 5505 CNRS, Toulouse, France
E-mail: guillaume.cabanac@univ-tlse3.fr, gilles.hubert@univ-tlse3.fr

B. Milard
Department of Sociology, University of Toulouse, LISST UMR 5193 CNRS, Toulouse, France
E-mail: beatrice.milard@univ-tlse2.fr

collaborations boost research outcomes: in terms of effect by a positive impact on citations received (Frenken, Hölzl, & de Vor, 2005; Wallace, Larivière, & Gingras, 2012) and in terms of structuring specialties by network's dynamic core (D. Lee, Goh, Kahng, & Kim, 2010). We wondered if these findings would also apply at the individual level of the single researcher.

S. Lee and Bozeman (2005) stressed the impact of collaboration on the increase in research productivity. The reason for this may be related with the theory of "preferential attachment" introduced by Barabási et al. (2002). Building on Price's (1976) concept of cumulative advantages, this theory states that some prominent researchers attract a large share of collaborators, turning these attractive researchers into hubs in the network of co-authorships (see also Cotta & Merelo, 2006; Huang, Zhuang, Li, & Giles, 2008).

Seminal work by Price and Gürsey (1975) and follow-up studies (e.g., Braun, Glänzel, & Schubert, 2001; Schubert & Glänzel, 1991) categorized authors according to their collaboration patterns. The core of collaboration networks is made up of so-called "continuants" around whom other types of researchers revolve, namely "newcomers," "terminators," and "transient" collaborators (i.e., ephemeral co-authors). Besides considering this variety of patterns as the mark of idiosyncratic preferences when collaborating, Price and Gürsey (1975) relate it to the demographics of scientists. The core of collaboration networks composed of continuants gather around the most productive researchers who happen to be senior researchers on the research scene. Nonetheless, there are also highly productive researchers who do not work with many collaborators (Braun et al., 2001). One might wonder, then, to what extent they continue to work with previous collaborators? Responses to this question suggest the need for further longitudinal and fine-grained studies of individual researchers.

This article investigates the dynamics of co-authorships from a career-wise perspective. We study a cohort of 3,860 researchers drawn from the scientific domain of Computer Science (CS) over 30 years. Our work extends previous research on individual researchers that suggested that collaborations result from local opportunities and collaborative arrangements elaborated with the passage of time (Cronin, Shaw, & Barre, 2004; Genuth, Chompalov, & Shrum, 2000; Sugimoto & Cronin, 2012). Here, we ask additional questions, such as: What are the scientific career profiles of co-authors involved throughout their careers? How do researchers reify fruitful collaborations in sustained co-authorships? Do researchers tend to maintain the same collaborations or, on the contrary, do they renew their co-authors regularly?

Data

This article relies on empirical observations of a cohort composed of 3,860 researchers. We first introduce the bibliographic dataset we used that records the research publications. We then report descriptive statistics on both 1) the population of researchers present in the whole dataset, and 2) the cohort of sampled researchers under study. These observations reveal features of the data regarding publication rates, and co-authorship intensities.

Demographics of the Population

The DBLP Computer Science Bibliography (Ley, 2002) records bibliographical information about worldwide research in computer science. It also indexes venues concerned with applied computing (e.g., biocomputing, scientometrics). An individual's bibliographic record lists the scientific output of a researcher during his/her career. The DBLP is both online,[1] and publicly

[1] `http://dblp.uni-trier.de/`

released as an XML file that we downloaded on March 11, 2012. This dataset records the bibliographies of 1,095,174 authors, totaling 1,919,594 documents. The quality of records is a key concern for DBLP (Reuther, Walter, Ley, Weber, & Klink, 2006). Non-western names are transliterated to the ISO-8859-1 (latin1) encoding. Name-related inconsistencies (e.g., homonyms, synonyms) are identified through network analysis of co-authorships, manual detection, and error reports sent by users (Reitz & Hoffmann, 2010, 2013). We refer the interested reader to the DBLP FAQ[2] for a detailed discussion of these name-related issues.

We focused on journal articles ($N = 784,661$) and paper conferences ($N = 1,085,393$), since these two kinds of materials have been shown to reflect research advances in CS (Chen & Konstan, 2010; Freyne, Coyle, Smyth, & Cunningham, 2010). Note that we discarded pre-prints from the ArXiv CoRR (Computing Research Repository), as these are not peer-reviewed materials and they may also repeat documents published in journals or conferences.

The longitudinal analysis of the DBLP shown in Figure 1 suggests few publications recorded from 1936 to 1980 (Period 1). There is a steady increase between 1980 and 2002 (Period 2), followed by a sharp increase between 2002 and 2010 (Period 3). The low number of papers recorded in the early years might result from Computer Science being a relatively new domain: The *Association for Computing Machinery* was founded in 1947. The increase in numbers from 1993 and onwards may coincide with the advent of the DBLP (Ley, 2002). Materials published before 1993 were entered by hand, whereas nowadays the system automatically extracts most of the bibliographic records. The DBLP covers a large share of research output in CS (Reitz & Hoffmann, 2010) and various studies have relied on it to reveal the features and publication culture of CS (e.g., see Cabanac, 2012, 2013; Cavero, Vela, & Cáceres, 2014; Deng, King, & Lyu, 2008; Elmacioglu & Lee, 2005; Solomon, 2009). The relative decrease in record numbers for year 2011 may result from the latency of data collection and processing.

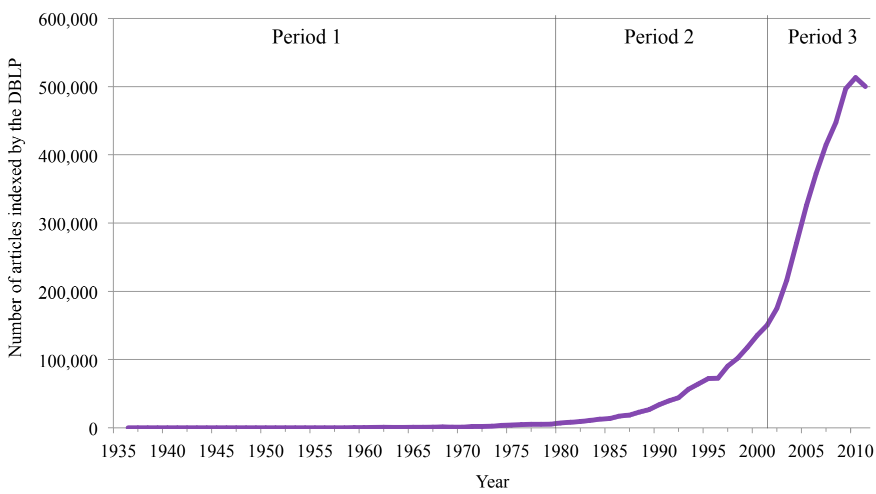

Fig. 1 Yearly research production of the population per year, in terms of conference papers and journal articles ($N = 1,870,054$), as recorded in the DBLP.

[2] `http://dblp.uni-trier.de/~ley/faq/Data+in+dblp.html`

The number of authors with x papers follows a Lotka (1926) distribution, as shown in Figure 2. This finding is in line with a previous observation on the 1968–2003 publications of a CS subfield (Elmacioglu & Lee, 2005). The head of the distribution (comprising the 99% of authors having published between 1 and 55 papers) contained more authors than we would expect according to the Lotka distribution. At the same time, one would expect to find more authors in the tail of the distribution.

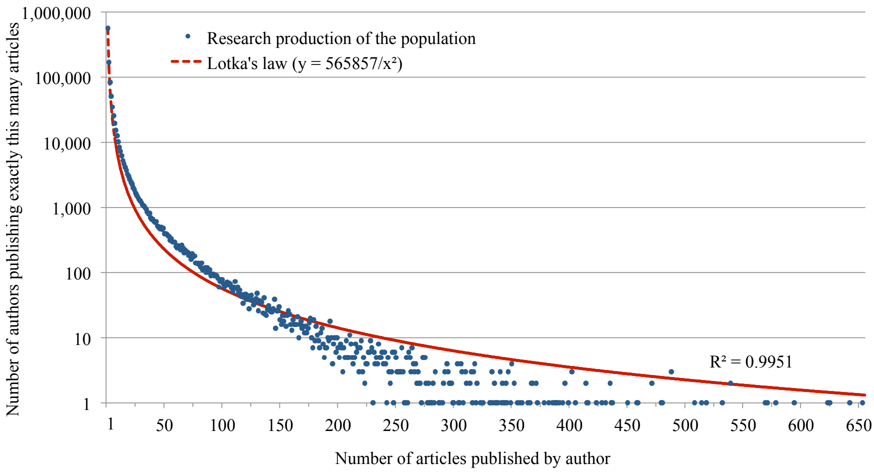

Fig. 2 Research production of the population of researchers, in terms of conference papers and journal articles ($N = 1,870,054$), as recorded in the DBLP. This distribution fits the Lotka's law ($R^2 = 0.9951$).

Demographics of the Sample under Study

This study seeks to uncover the features of academic careers in CS. We thus sampled researchers from the DBLP according to various criteria regarding seniority and productivity. First, we sampled researchers who began publishing between 1980 and 1985, with their last publication appearing in 2005 or later. These requirements intend to sample senior researchers with an active 21-year-long career (1985–2005) to 32-year-long career (1980–2011). Second, we discarded those researchers with less than 15 published papers throughout their careers. This requirement allowed us to focus on researchers with a steady research output throughout their careers. Eventually, we arrived at a cohort of 3,860 researchers (denoted $\mathcal{R}$) having published a total of 209,377 journal articles and conference papers (denoted $\mathcal{P}$). In the remainder of this article, we refer to a subject of the cohort as a "fifty-something researcher."

The yearly publication output of the cohort (Figure 3) increased linearly between 1980 and 2005. Then, it plateaued in 2006–2008 before decreasing noticeably in 2009–2011. Similar phenomena have been extensively discussed nationwide (Costas, van Leeuwen, & Bordons, 2010; Gingras, Larivière, Macaluso, & Robitaille, 2008; Kyvik & Olsen, 2008) or at the level of particular scientific domains, such as psychology (Over, 1982), physics, earth science, physiology, and biochemistry (Levin & Stephan, 1989), and physics again (Hall, Mairesse, & Turner, 2007). Various factors of productivity decline with time were discussed in the literature (e.g., see Goodwin & Sauer, 1995; Kyvik & Olsen, 2008; Stroebe, 2010). We refer the interested reader to (Kyvik & Olsen, 2008) for a review of these factors grouped into six hypotheses: The utility maximizing hypothesis, the seniority burden hypothesis, the cumulative disadvantage hypothesis, the age decrement hypothesis, the obsolescence hypothesis, and the intellectual deadlock hypothesis.

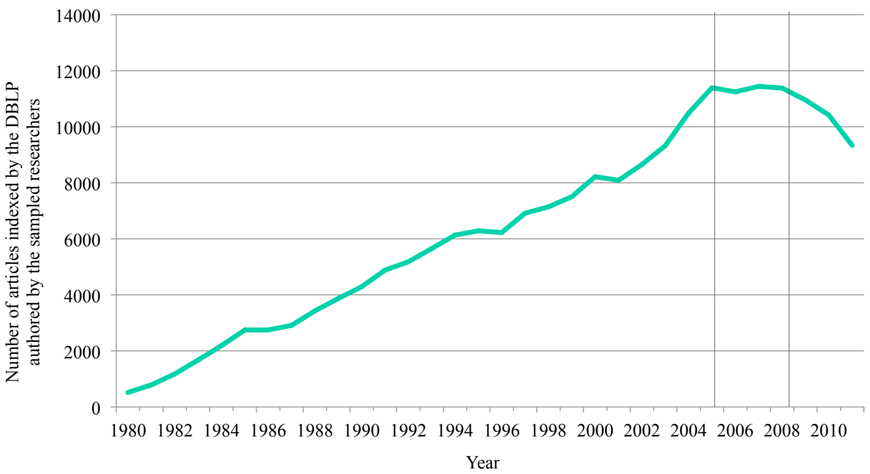

Fig. 3 Yearly research production of the sampled 3,860 researchers, in terms of conference papers and journal articles ($N = 209,377$), as recorded in the DBLP.

Our cohort's production shown in Figure 4 fits the Lotka's distribution, but to a lesser extent than did the population (Figure 2). We note, however, that the cohort's production distribution is underestimated by Lotka's law. This suggests that our sampled researchers outperformed the production expected from regular researchers. This can be explained by their longevity (about 25 years) and their productivity (at least 15 papers) on the academic scene.

Our idea in this article is that the relational structure of co-authorships can also suggest new hypotheses about these co-authorships and their evolution. We address these hypotheses with a transversal and a longitudinal studies in the following sections.

Transversal study of academic careers

We first studied the co-authorships of papers published by the 3,860 researchers in our study. The transversal study of the entire career of the researchers in $\mathcal{R}$ reveals that 50% of them had between 11 and 34 co-authors ($Mdn = 20$), as shown by the box of the boxplot in Figure 5. Overall, 95% of the fifty-somethings had between 0 and 68 co-authors, as shown by the whiskers. Notice too that only 30 researchers (0.8%) had no collaborators at all. The remaining 5% of outlying researchers (black dots) had more than 68 co-authors.

The boxplot in Figure 6 shows the partnership ability φ-index (Schubert, 2012) of the sampled researchers. The median $\varphi = 4$, which suggests that a typical author has published 4 papers with 4 of his/her co-authors each. Notice the high variance among researchers, with outliers above $\varphi = 10$ showing hyper-collaborative researchers.

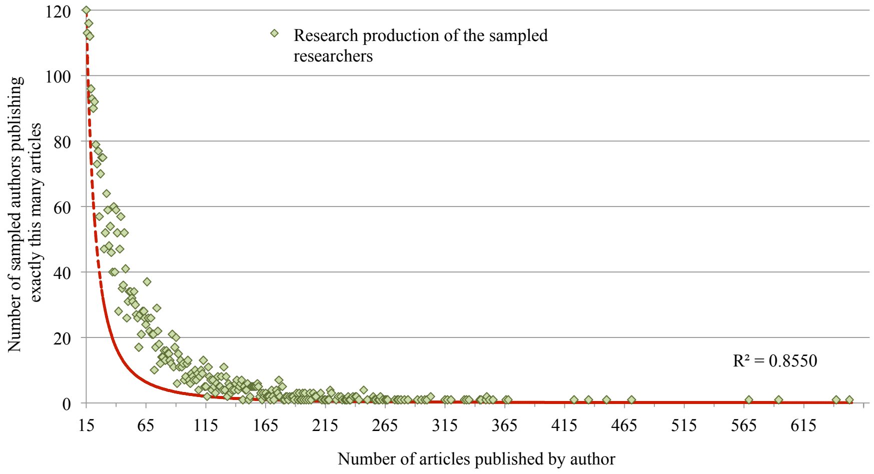

Fig. 4 Research production of the 3,860 sampled researchers, in terms of conference papers and journal articles ($N = 209,377$), as recorded in the DBLP. This distribution fits the Lotka's law ($R^2 = 0.8550$).

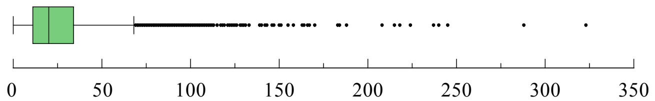

Fig. 5 Average number of co-authors for the researchers under study ($N = 3,860$).

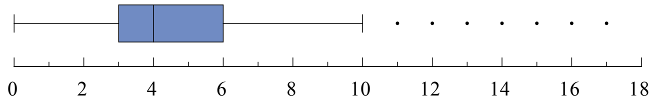

Fig. 6 Partnership ability φ-index of the 3,860 researchers under study showing the distribution of the number of φ co-authors per researcher.

We computed the share of transient co-authors for the 3,830 researchers under study with collaborators (Figure 7). We found that 50% of $\mathcal{R}$ did not pursue their publication collaboration with between 62.5% and 83.9% of their new co-authors ($Mdn = 72.7\%$), as shown by the box of the boxplot in Figure 7. Overall, 98,6% of the fifty-somethings have abandoned their collaboration with between 30.6% and 100% of their new co-authors, as shown by the whiskers. The remaining 1.4% of outlying researchers secured more than 69.4% of their new co-authors, as shown by the black dots. Figure 7 suggests that most of the researchers did not collaborate again with at least two-third of their co-authors.

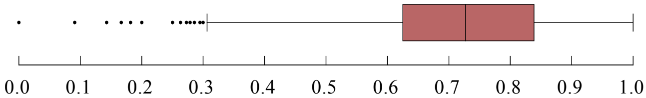

Fig. 7 Average share of transient co-authors per researcher under study who had collaborators ($N = 3,830$).

One may wonder about the nature of these ephemeral collaborations. We illustrate this point with a case study on the co-authorship of an active fifty-somethings in the $\mathcal{R}$

sample: Dr. Serge Abiteboul, senior researcher, INRIA, France. Abiteboul has 238 referred papers in the $\mathcal{P}$ sample with a total of 200 co-authors. Transient co-authors represent 50% ($N = 99$) of Abiteboul's co-authors, with whom he co-signed 16% ($N = 40$) of his papers. These unbalanced figures imply that some papers were co-authored with several transient researchers.

We studied Abiteboul's ephemeral collaborations by browsing the 40 aforementioned papers with a special attention to the 'acknowledgment' sections. This allowed us to determine the following scientific contexts:

- *Reporting of Current Research: 60.6% of transient co-authors ($N = 60$).* This category refers to papers reporting research in collaboration with other researchers of variable seniority. Most of these collaborations occur in the context of national and international projects gathering multiple partners ($N = 34$). For example, (Pitoura, Abiteboul, Pfoser, Samaras, & Vazirgiannis, 2003) gathers 5 researchers involved in the DBGlobe European project, 7 of which are part of Abiteboul's transient co-authors. Other collaborations (e.g., Abiteboul, Alstrup, Kaplan, Milo, & Rauhe, 2006) result from the one-time collaboration of two research teams tackling a common research issue ($N = 18$). Note that only two PhD students supervised by Abiteboul are transient co-authors, since the defending of a PhD in Computer Science generally requires that the author published his work in conference of journal papers beforehand. Eventually, few papers ($N = 5$) were written with transient visiting scholars (e.g., Abiteboul & Bonner, 1991).

- *Promotion of Research Directions: 26.3% of transient co-authors ($N = 26$).* This category refers to papers stemming from working groups gathering senior researchers. Such papers review the research literature of the considered domain (i.e., databases) and stress research directions. For example, Abiteboul coauthored papers with the Lowell Database Group (Abiteboul, Gawlick, et al., 2005), with the DELOS-NSF working group (Ioannidis et al., 2005), and in the context of conference panels and various meetings (e.g., Özsu et al., 2011).

- *Celebration of Past Research: 13.1% of transient co-authors ($N = 13$).* This category refers to papers/obituaries honoring the memory of influential researchers. Each of the co-authors commemorates the late colleague by writing a piece of text that is included in the paper. For instance, Abiteboul, Hull, and Vianu (2005) honor the memory of Professor Seymour Ginsburg. In reminiscence-type papers, each co-author revisit and discuss past landmark papers. For example, ten senior researchers discuss the paper that most influenced their research in (Snodgrass et al., 1999). In this very paper, Abiteboul commented a landmark paper published 19 years before.

Given the large proportion of transient collaborators, one may wonder how many recurrent co-authors (i.e., non transient) are really involved in an author's career? Overall, 50% of the researchers had managed to secure between 2 and 11 recurrent co-authors ($Mdn = 5$), as shown by the box of the boxplot in Figure 8. Overall, 95% of the fifty-somethings had between 0 and 25 co-authors, as shown by the whiskers. The remaining 6% of outlying researchers (black dots) had more than 25 co-authors.

Summing up, the lifelong collaborations of the 3,860 researchers mainly involved transient collaborators ($Mdn = 72.7\%$, see Figure 7) and a median of 5 recurrent co-authors (Figure 8). The longitudinal study of the data reveals further findings discussed in the following section.

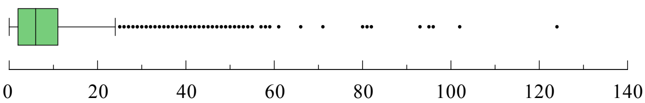

Fig. 8 Average number of recurrent co-authors for the researchers under study ($N = 3,860$).

Longitudinal study of academic careers

We wondered whether researchers kept working with the same people over and over again, or if they frequently renewed their co-authorships. Of course, there may be several intermediate and less extreme collaboration patterns. To answer this question, we had to distinguish between brand-new and prior collaborations involved in each paper of $\mathcal{P}$. We describe below the method we set up, and the results of the associated longitudinal study.

Method

For any given paper $p \in \mathcal{P}$ published by an author $a \in \mathcal{R}$, we considered the n co-authors $\{c_1, \dots, c_i, \dots, c_n\}$ of a. We classified any co-author c_i in one of the two classes: "New" or "Former" co-author. The assignment of any c_i to its class depends on whether we detect any previously published paper p' (i.e., $year(p') < year(p)$) jointly written by a and c_i ("former co-author") or not ("new co-author").

Let us illustrate this process with the example of John Smith, one of the fifty-something researchers under study (i.e., in $\mathcal{R}$). For each paper published by John Smith (author a), we consider all his authors c_i. Table 1 shows an example of a paper published in 2009 with five co-authors: [John Doe, Hélène Haztaquès, Ike Antkare, *John Smith*, Ashok Kumar, Erika Mustermann]. The function $prevCollab(a, c_i, y) \in \mathbb{N}^+$ computes the number n of previous collaborations between a (John Smith) and c_i, which happened before year y. This is used to assign each c_i to a class. The value $n > 0$ shows former collaborators of John Smith as of year y, whereas $n = 0$ shows new collaborators.

Table 1 Data about the co-authors of John Smith (a) in an example article published in year $y = 2009$

Co-author c_i	Previous Collaborations		Expertise of c_i	
	$prevCollab(a, c_i, y)$	Label	$prevPapers(c_i, y)$	Label
John Doe	0	New	51	Confirmed
Hélène Haztaquès	3	Former	8	Intermediate
Ike Antkare	0	New	2	Intermediate
Ashok Kumar	17	Former	24	Confirmed
Erika Mustermann	0	New	0	Newcomer

Besides identifying former collaborations, we also wondered about the experience in publishing of each co-author c_i at the time of the collaboration (i.e., publication of the paper). The function $prevPapers(c_i, y) \in \mathbb{N}^+$ computes the number n' of papers published by c_i before year y. This production indicator is used to classify each c_i in one class, namely "Newcomer" ($n' = 0$), "Intermediate" ($n' > 0$) or "Confirmed" ($n' > 0$ as well).

Instead of intuitively and arbitrarily defining the boundary between "Intermediate" and "Confirmed" classes, we assumed that the 10% of most prolific authors would represent confirmed researchers. For example, in 2009, the 10th percentile of papers per author was 9. As a result, researchers having published between 1 and 8 papers before 2009 are denoted as "Intermediate" for year 2009. Note that this threshold was computed for each year, as publication habits (and average production) have evolved over time. The globalization of science and publish-or-perish atmosphere (Garfield, 1996) may be factors in the steady increase of this threshold (from 4 papers in 1980 to 9 papers in 2011).

Results

We studied the sample of papers produced by the 3,860 researchers in our study. Figure 9 shows the evolution of the average number of co-authors per paper. We found a strong linear relationship ($R^2 = 0.9411$) between time and the average number of co-authors per paper, which has increased steadily since the late 1980's. Sampled researchers authored papers with an average of 1.8 colleagues at the beginning of their career (1980), while they collaborated with about 3 colleagues per paper during the evening of their careers (2011).

In the following sections, we study the evolution of co-authorships during the career of the 3,860 sampled researchers between 1980 and 2011. We focus on 1) the transience vs. continuance of collaborators, and 2) on their expertise.

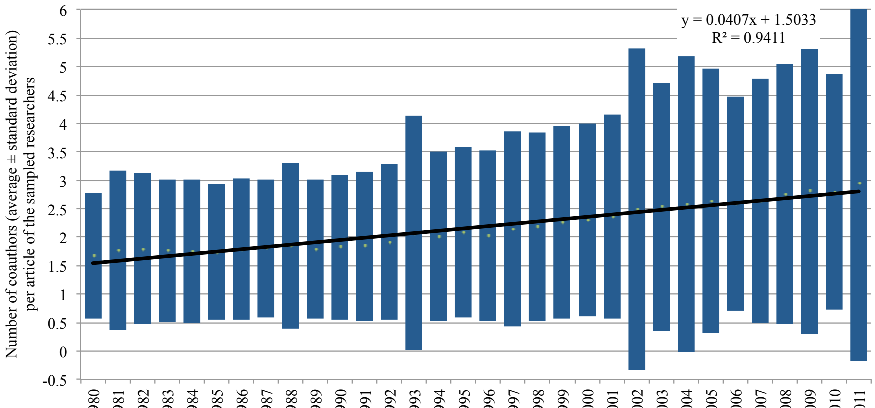

Fig. 9 Evolution of the number of co-authors per article showing an increase during the period under study.

Collaborations through a researcher's career: Transient vs. recurrent collaborators

We have already pointed to the increasing number of authors per paper (Figure 9). In this section, we study the distribution of *former* and *new* co-authors in time. Figure 10 shows that throughout their careers, researchers kept on introducing new collaborators. Meanwhile, the share of former co-authors per paper kept on increasing as time passed. One may thus distinguish three stages during the careers under study:

a. 1980–1985: the *early* stage saw a large number of new co-authors per paper, with few former co-authors (as expected since the researchers began their career at this stage).

b. 1986–1993: the *mid* stage saw a balanced number of new vs. former co-authors per paper, with an overall of 2 co-authors per paper.
c. 1994–2011: the *late* stage saw a larger number of former co-authors than new co-authors.

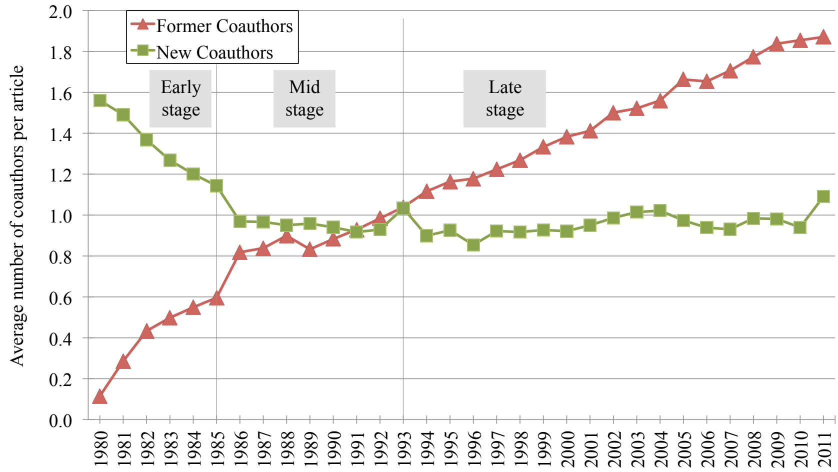

Fig. 10 Evolution of the balance between Former and New co-authors for the 3,860 researchers under study.

According to Figure 10, however, we cannot determine whether a) ties with new co-authors are consolidated through further collaborations (i.e., do new co-authors become former co-authors?), or b) new co-authors are transient, one-shot partners who do not integrate the group of Former co-authors afterwards. Our transversal study of academic careers (Figure 7) supports the latter option, since a median of 72.7% of all co-authors are transient, one-shot partners.

Expertise of co-authors

Next, we wondered about the expertise of the collaborators. According to the theory of Barabási et al. (2002) about preferential attachments, one would expect that the researchers under study would be increasingly attractive because of their growing recognition in the scientific community. Figure 11 shows the distribution of collaborators according to their expertise longitudinally. It must be remembered that the researchers under study themselves were Newcomers at the beginning the 1980's, then became Intermediate, and finally Confirmed, since they had all published at least 15 publications (i.e., more than the aforementioned 10th percentile of paper production per capita). We found that the number of Newcomers slightly decreases as the time passes (Figure 11). Nevertheless, researchers never stopped working with Newcomers as co-authors throughout their career. As our researchers gained in experience, they progressively worked with more Intermediate and Confirmed co-authors. This observation may be related to homophily in academia (Kossinets & Watts, 2009), which is best illustrated by the saying "birds of a feather flock together."

A final question remains: What is the expertise of the co-authors attracted by the researchers under study, when considering New and Former coauthors apart? Refining Figures 10 and 11, we show in Figure 12 the expertise of co-authors (i.e., Newcomer, Intermediate, Confirmed) according to their authorship status, namely New or Former co-authors.

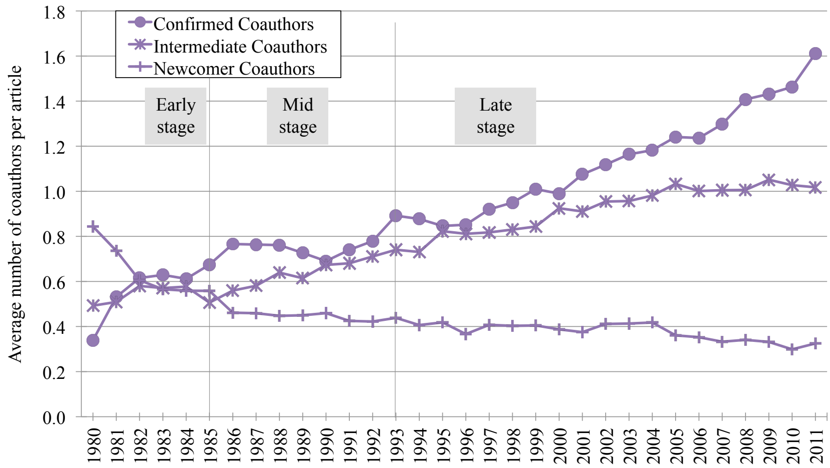

Fig. 11 Evolution of the balance between Newcomer, Intermediate, and Confirmed co-authors for the 3,860 researchers under study.

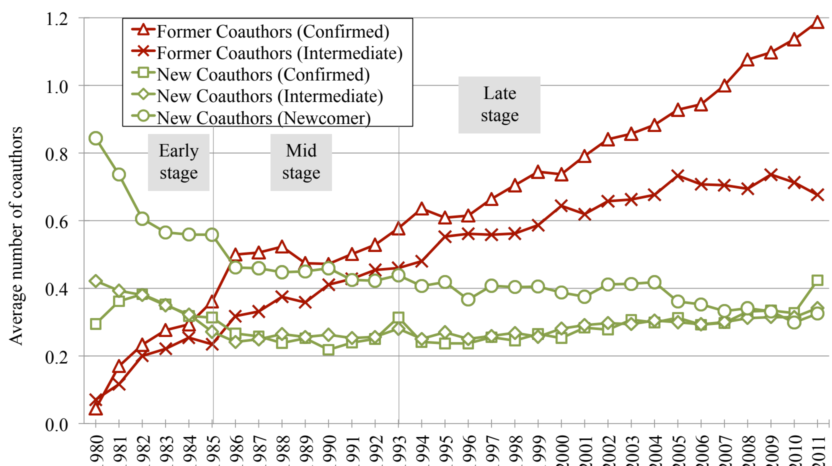

Fig. 12 Evolution of the expertise balance (Newcomer, Intermediate, and Confirmed) of co-authors for the Former and New co-author categories for the 3,860 researchers under study.

During the early stage, most of the new co-authors were unexperienced (Newcomers) and Former co-authors were Confirmed. This may reflect the relationship of a researcher with colleagues and his/her research supervisors. During the mid stage, there is an increase in the number of Former co-authors, with the share of Confirmed scientists growing. This increase continued during the late stage. Notice that researchers succeeded in attracting one-third of a "New and Confirmed" author per paper, and one-third of a "New and Intermediate" author per paper. All in all, New co-authors are equally distributed among the three classes of expertise. Scientists engage new collaborations with researchers with a full range of expertise. The growth of co-authors per paper stems from the Confirmed group of Former co-authors, especially during the late stage of the career (see the triangles on the top-right corner of Figure 12).

Discussion

Researchers develop their careers idiosyncratically with respect to the topics that they address, as recently discussed in (Horlings & Gurney, 2013). In this paper, we also stress the influence of social dynamics on the shaping of careers.

The study of co-authorships is not novel *per se* (e.g., see Glänzel & Schubert, 2005). At an aggregated level, they provide hints of the transience and continuance of collaborations at a national level (Cheng, Hen, Tan, & Fok, 2013; de Souza & Azevedo Ferreira, 2013; Schubert & Glänzel, 1991), or on the construction and evolution of subfields inside a research domain (D. Lee et al., 2010; Velden, Haque, & Lagoze, 2010). At the individual level, they provide hints on researcher careers as co-authorships convey the core of each researcher's activity. For example, Zhang and Glänzel (2012) analyzed the co-authorships of four eminent researchers in different disciplines to show discrepancies. Our study reveals the average tendencies in researchers' co-authorships in pure and applied Computer Science. We found that researchers increasingly worked with Former co-authors in new publications as their careers move on. The attractiveness of New "Confirmed" (i.e., expert) co-authors, although being slightly more important at the beginning of researchers' careers then rapidly levels off throughout researchers' careers. Co-authors held in esteem are familiar collaborators with whom they repeatedly collaborated. This suggests that preferential attachments are probably supported by social phenomena, such as homophily and acquaintanceship.

Our study helps to lay the groundwork for a careful analysis of small groups formed by co-authors of any article. Several extensions to this type of research can be suggested in order to deepen some of the current issues involving scientific collaboration. Studies of small groups of co-authors would allow the reexamination of the social structure of the domain under study, as a complement to studies at an aggregated level. In the same way, we would have to study the "lost" co-authors (i.e., terminators) of our researcher cohort: Is there a period in a researcher's career when he/she no longer publishes with co-authors with whom he/she extensively published previously? To what extent is it reciprocated? Interviews with researchers drawn from the cohort would allow us to understand better these cases of "lost" co-authors. Furthermore, understanding long partnerships would also be interesting: What types of relations lead to such partnerships (e.g., subject matters and/or social proximity through institution, friendship or family bonds)? Are there negative relationships affecting the careers of researchers (e.g., "sticky" or exclusive co-authors preventing one from working with other colleagues)? Such questions require qualitative study, as an essential complement to the quantitative research presented in this article.

Summary and Conclusion

This research focused on the study of collaborations among researchers. Our approach is different from numerous studies on co-authorship networks, as it relies on a longitudinal study of lifelong careers of a researcher cohort. We mined the publication records of this cohort extracted from the DBLP Bibliography related to Computer Science (extracted from article bylines). We considered 3,860 researchers who published at least 15 papers in Computer Science conferences and journals alike, provided that they started publishing between 1980 and 1985, and were still active from 2005 onward.

Our study revealed that the 3,860 researchers collaborated with a sustained number of new co-authors throughout their careers. A median of 72.7% co-authors were transient and the circle of recurrent co-authors comprised a median of 5 collaborators. We found that the number of newcomers among co-authors slightly decreased with the passage of time. Nevertheless, researchers never stopped working with newcomers as co-authors throughout their careers. As researchers gained in experience, they progressively worked with more intermediate and confirmed co-authors, with whom they already published. New co-authors were equally distributed among the three classes of expertise (newcomers, intermediate, and confirmed).

We see several extensions to this type of research to deepen some current issues in the study of scientific collaborations. Studying the hierarchies within small groups of co-authors might reveal other forms of research stratification, and it might give a new slant on the old issue of inequalities in research (J. R. Cole & Zuckerman, 1984; S. Cole, 1979). One extension of our work will be to make a differentiated analysis of sub-cohorts of researchers to better understand their characteristics (according to gender, geographical origin, institutional origins, and so on) and to further elaborate the structural and demographic constraints that influence potential collaboration.

Acknowledgements This work was supported by the French National Agency for Research (ANR-11-BSH1-0013). We acknowledge the feedback of James Hartley and András Schubert on an earlier version of this article.

References

Abiteboul, S., Alstrup, S., Kaplan, H., Milo, T., & Rauhe, T. (2006). Compact labeling scheme for ancestor queries. *SIAM Journal on Computing*, *35*(6), 1295–1309. doi: 10.1137/s0097539703437211

Abiteboul, S., & Bonner, A. (1991). Objects and views. *SIGMOD Record*, *20*(2), 238–247. doi: 10.1145/119995.115830

Abiteboul, S., Gawlick, D., Gray, J., Haas, L., Halevy, A., Hellerstein, J., … Molina, H. G. (2005). The Lowell database research self-assessment. *Communications of the ACM*, *48*(5), 111–118. doi: 10.1145/1060710.1060718

Abiteboul, S., Hull, R., & Vianu, V. (2005). In memory of Seymour Ginsburg 1928–2004. *SIGMOD Record*, *34*(1), 5–12. doi: 10.1145/1058150.1058152

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2009). Research collaboration and productivity: Is there correlation? *Higher Education*, *57*(2), 155–171. doi: 10.1007/s10734-008-9139-z

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*(3–4), 590–614. doi: 10.1016/S0378-4371(02)00736-7

Beaver, D., & Rosen, R. (1978). Studies in scientific collaboration. *Scientometrics*, *1*(1), 65–84. doi: 10.1007/BF02016840

Braun, T., Glänzel, W., & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics*, *50*(3), 499–510. doi: 10.1023/A:1012774206340

Cabanac, G. (2012). Shaping the landscape of research in information systems from the perspective of editorial boards: A scientometric study of 77 leading journals. *Journal of the American Society for Information Science and Technology*, *63*(5), 977–996. doi: 10.1002/asi.22609

Cabanac, G. (2013). Experimenting with the partnership ability φ-index on a million computer scientists. *Scientometrics*, *96*(1), 1–9. doi: 10.1007/s11192-012-0862-y

Cavero, J. M., Vela, B., & Cáceres, P. (2014). Computer science research: More production, less productivity. *Scientometrics*, *98*(3), 2103–2111. doi: 10.1007/s11192-013-1178-2

Chen, J., & Konstan, J. A. (2010). Conference paper selectivity and impact. *Communications of the ACM*, *53*(6), 79–83. doi: 10.1145/1743546.1743569

Cheng, M. Y., Hen, K. W., Tan, H. P., & Fok, K. F. (2013). Patterns of co-authorship and research collaboration in Malaysia. *Aslib Proceedings*, *65*(6), 659–674. doi: 10.1108/AP-12-2012-0094

Cole, J. R., & Zuckerman, H. (1984). The productivity puzzle: Persistence and change in patterns of publication of men and women scientists. In M. W. Steinkempt & M. L. Maehr (Eds.), *Advances in motivation and achievement* (pp. 217–258). Greenwich, Conn.: JAI Press.

Cole, S. (1979). Age and scientific performance. *American Journal of Sociology*, *84*(4), 958–977.

Costas, R., van Leeuwen, T. N., & Bordons, M. (2010). A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity and impact. *Journal of the American Society for Information Science and Technology*, *61*(8), 1564–1581. doi: 10.1002/asi.21348

Cotta, C., & Merelo, J.-J. (2006). The complex network of EC authors. *SIGEVOlution*, *1*(2), 2–9. doi: 10.1145/1147192.1147193

Cronin, B., Shaw, D., & Barre, K. L. (2004). Visible, less visible, and invisible work: Patterns of collaboration in 20th century chemistry. *Journal of the American Society for Information Science and Technology*, *55*(2), 160–168. doi: 10.1002/asi.10353

Deng, H., King, I., & Lyu, M. R. (2008). Formal Models for Expert Finding on DBLP Bibliography Data. In *ICDM'08: Proceedings of the 8th IEEE International Conference on Data Mining* (pp. 163–172). IEEE Computer Society. doi: 10.1109/ICDM.2008.29

de Souza, C. G., & Azevedo Ferreira, M. L. (2013). Researchers profile, co-authorship pattern and knowledge organization in Information Science in Brazil. *Scientometrics*, *95*(2), 673–687. doi: 10.1007/s11192-012-0882-7

Elmacioglu, E., & Lee, D. (2005). On six degrees of separation in DBLP-DB and more. *SIGMOD Record*, *34*(2), 33–40. doi: 10.1145/1083784.1083791

Evans, T. S., Lambiotte, R., & Panzarasa, P. (2011). Community structure and patterns of scientific collaboration in business and management. *Scientometrics*, *89*(1), 381–396. doi: 10.1007/s11192-011-0439-1

Frenken, K., Hölzl, W., & de Vor, F. (2005). The citation impact of research collaborations: The case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering and Technology Management*, *22*(1–2), 9-30. doi: 10.1016/j.jengtecman.2004.11.002

Freyne, J., Coyle, L., Smyth, B., & Cunningham, P. (2010). Relative status of journal and conference publications in Computer Science. *Communications of the ACM*, *53*(11), 124–132. doi: 10.1145/1839676.1839701

Garfield, E. (1996). What is the primordial reference for the phrase 'Publish or Perish'? *The Scientist*, *10*(12), 11.

Genuth, J., Chompalov, I., & Shrum, W. (2000). How experiments begin: The formation of scientific collaborations. *Minerva*, *38*(3), 311–348. doi: 10.1023/A:1026573717027

Gingras, Y., Larivière, V., Macaluso, B., & Robitaille, J.-P. (2008). The effects of aging on researchers publication and citation patterns. *PLoS ONE*, *3*(12), e4048. doi: 10.1371/journal.pone.0004048

Glänzel, W., & Schubert, A. (2005). Analysing scientific networks through co-authorship. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 257–276). Springer. doi: 10.1007/1-4020-2755-9_12

Goodwin, T. H., & Sauer, R. D. (1995). Life cycle productivity in academic research: Evidence from cumulative publication histories of academic economists. *Southern Economic Journal*, *61*(3), 728–743. doi: 10.2307/1060993

Hall, B. H., Mairesse, J., & Turner, L. (2007). Identifying age, cohort, and period effects in scientific research productivity: Discussion and illustration using simulated and actual data on French physicists. *Economics of Innovation and New Technology*, *16*(2), 159–177. doi: 10.1080/10438590600983010

Horlings, E., & Gurney, T. (2013). Search strategies along the academic lifecycle. *Scientometrics*, *94*(3), 1137-1160. doi: 10.1007/s11192-012-0789-3

Huang, J., Zhuang, Z., Li, J., & Giles, C. L. (2008). Collaboration over time: Characterizing and modeling network evolution. In *WSDM'08: Proceedings of the international conference on web search and web data mining* (pp. 107–116). New York, NY, USA: ACM. doi: 10.1145/1341531.1341548

Ioannidis, Y., Maier, D., Abiteboul, S., Buneman, P., Davidson, S., Fox, E., … Weikum, G. (2005). Digital library information-technology infrastructures. *International Journal on Digital Libraries*, *5*(4), 266-–274. doi: 10.1007/s00799-004-0094-8

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18. doi: 10.1016/S0048-7333(96)00917-1

Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, *115*(2), 405–450. doi: 10.1086/599247

Kyvik, S., & Olsen, T. B. (2008). Does the aging of tenured academic staff affect the research performance of universities? *Scientometrics*, *76*(3), 439–455. doi: 10.1007/s11192-007-1767-z

Landry, R., & Amara, N. (1998). The impact of transaction costs on the institutional structuration of collaborative academic research. *Research Policy*, *27*(9), 901–913. doi: 10.1016/S0048-7333(98)00098-5

Lee, D., Goh, K.-I., Kahng, B., & Kim, D. (2010). Complete trails of coauthorship network evolution. *Physical Review E*, *82*(2), 026112-1–026112-9. doi: 10.1103/PhysRevE.82.026112

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673–702. doi: 10.1177/0306312705052359

Levin, S. G., & Stephan, P. E. (1989). Age and research productivity of academic scientists. *Research in Higher Education*, *30*(5), 531–549. doi: 10.1007/BF00992202

Ley, M. (2002). The DBLP computer science bibliography: Evolution, research issues, perspectives. In A. H. F. Laender & A. L. Oliveira (Eds.), *SPIRE'02 : Proceedings of the 9th international conference on String Processing and Information Retrieval* (Vol. 2476, pp. 1–10). Springer. doi: 10.1007/3-540-45735-6_1

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*(12), 317–324.

Over, R. (1982). Does research productivity decline with age? *Higher Education*, *11*(5), 511–520. doi: 10.1007/BF00194416

Özsu, M. T., Valduriez, P., Abiteboul, S., Kemme, B., Jiménez-Péris, R., & Ooi, B. C. (2011). Distributed data management in 2020? *ICDE'11 Proceedings of the 27th IEEE International Conference on Data Engineering*. doi: 10.1109/icde.2011.5767962

Pitoura, E., Abiteboul, S., Pfoser, D., Samaras, G., & Vazirgiannis, M. (2003). DBGlobe: A service-oriented P2P system for global computing. *SIGMOD Record*, *32*(3), 77–82. doi: 10.1145/945721.945737

Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306. doi: 10.1002/asi.4630270505

Price, D. d. S., & Gürsey, S. (1975). Studies in scientometrics I: Transience and continuance in scientific authorship. *Ciência da Informação*, *4*(1), 27–40.

Reitz, F., & Hoffmann, O. (2010). An analysis of the evolving coverage of computer science sub-fields in the DBLP digital library. In M. Lalmas, J. Jose, A. Rauber, F. Sebastiani, & I. Frommholz (Eds.), *ECDL'10: Proceedings of the 14th European Conference on Research and Advanced Technology for Digital Libraries* (Vol. 6273, p. 216-227). Springer. doi: 10.1007/978-3-642-15464-5_23

Reitz, F., & Hoffmann, O. (2013). Learning from the past: An analysis of person name corrections in the DBLP collection and social network properties of affected entities. In T. Özyer, J. Rokne, G. Wagner, & A. H. Reuser (Eds.), *The Influence of Technology on Social Network Analysis and Mining* (Vol. 6, pp. 427–453). Springer. doi: 10.1007/978-3-7091-1346-2_19

Reuther, P., Walter, B., Ley, M., Weber, A., & Klink, S. (2006). Managing the quality of person names in DBLP. In J. Gonzalo, C. Thanos, M. F. Verdejo, & R. C. Carrasco (Eds.), *ECDL'06: Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries* (Vol. 4172, pp. 508–511). Springer. doi: 10.1007/11863878_55

Schubert, A. (2012). A Hirsch-type index of co-author partnership ability. *Scientometrics*, *91*(1), 303–308. doi: 10.1007/s11192-011-0559-7

Schubert, A., & Glänzel, W. (1991). Publication dynamics: Models and indicators. *Scientometrics*, *20*(1), 317–331. doi: 10.1007/BF02018161

Snodgrass, R. T., Abiteboul, S., Cluet, S., Franklin, M. J., Lohman, G. M., Lomet, D. B., … Valduriez, P. (1999). Reminiscences on influential papers. *SIGMOD Record*, *28*(1), 110–114.

Solomon, J. (2009). Programmers, professors, and parasites: Credit and co-authorship in Computer Science. *Science and Engineering Ethics*, *15*(4), 467–489. doi: 10.1007/s11948-009-9119-4

Stroebe, W. (2010). The graying of academia: Will it reduce scientific productivity? *American Psychologist*, *65*(7), 660–673. doi: 10.1037/a0021086

Stvilia, B., Hinnant, C. C., Schindler, K., Worrall, A., Burnett, G., Burnett, K., … Marty, P. F. (2011). Composition of scientific teams and publication productivity at a national science lab. *Journal of the American Society for Information Science and Technology*, *62*(2), 270–283. doi: 10.1002/asi.21464

Sugimoto, C. R., & Cronin, B. (2012). Biobibliometric profiling: An examination of multifaceted approaches to scholarship. *Journal of the American Society for Information Science and Technology*, *63*(3), 450–468. doi: 10.1002/asi.21695

Velden, T., Haque, A.-u., & Lagoze, C. (2010). A new approach to analyzing patterns of collaboration in co-authorship networks: Mesoscopic analysis and interpretation. *Scientometrics*, *85*(1), 219–242. doi: 10.1007/s11192-010-0224-6

Wallace, M. L., Larivière, V., & Gingras, Y. (2012). A small world of citations? The influence of collaboration networks on citation practices. *PLoS ONE*, *7*(3), e33339. doi: 10.1371/journal.pone.0033339

Zhang, L., & Glänzel, W. (2012). Where demographics meets scientometrics: Towards a dynamic career analysis. *Scientometrics*, *91*(2), 617–630. doi: 10.1007/s11192-011-0590-8